6.004 Quiz 2 Review Problems

RISC-V Assembly instructions and their encoding:

<table>
<thead>
<tr>
<th>RV32I Base Instruction Set (MIT 6.S084 subset)</th>
</tr>
</thead>
<tbody>
<tr>
<td>imm[31:12]</td>
</tr>
<tr>
<td>imm[20:11:11:10:12]</td>
</tr>
<tr>
<td>imm[11:0]</td>
</tr>
<tr>
<td>imm[11:0]</td>
</tr>
<tr>
<td>imm[11:0]</td>
</tr>
<tr>
<td>imm[11:0]</td>
</tr>
<tr>
<td>imm[11:0]</td>
</tr>
<tr>
<td>imm[11:0]</td>
</tr>
<tr>
<td>imm[11:0]</td>
</tr>
<tr>
<td>imm[11:0]</td>
</tr>
<tr>
<td>0000000</td>
</tr>
<tr>
<td>0000000</td>
</tr>
<tr>
<td>0100000</td>
</tr>
<tr>
<td>0000000</td>
</tr>
<tr>
<td>0100000</td>
</tr>
<tr>
<td>0000000</td>
</tr>
<tr>
<td>0000000</td>
</tr>
<tr>
<td>0000000</td>
</tr>
<tr>
<td>0000000</td>
</tr>
<tr>
<td>0000000</td>
</tr>
<tr>
<td>0100000</td>
</tr>
<tr>
<td>0000000</td>
</tr>
<tr>
<td>0000000</td>
</tr>
</tbody>
</table>
Problem 1. RISC-V Assembly

A) What is the binary encoding of the following RISC-V instruction?

\[
\text{sub x1, x2, x3}
\]

B) For each of the following sequences of instructions, assume all registers are initially 0. Specify the value of each used register after executing the sequence of instructions. Specify the address corresponding to all labels. Also, specify the address and contents of any modified memory locations.

A)
\[
\begin{align*}
\text{li x0, 3} \\
\text{li x1, 5} \\
\text{lui x2, 6}
\end{align*}
\]

B)
\[
\begin{align*}
\text{li x4, 7} \\
\text{addi x5, x4, -2} \\
\text{xor x4, x5, x4}
\end{align*}
\]

C)
\[
\begin{align*}
\text{. = 0} \\
\text{addi x5, x5, 2} \\
\text{blt x4, x5, L1} \\
\text{addi x4, x4, 5} \\
\text{L1: addi x4, x4, -3}
\end{align*}
\]

D)
\[
\begin{align*}
\text{. = 0} \\
\text{li x2, 0x100} \\
\text{addi x6, x2, 4} \\
\text{lw x3, 8(x2)} \\
\text{li x4, 0x114} \\
\text{mv x5, x4} \\
\text{sw x4, 0xC(x6)}
\end{align*}
\]

\[. = 0x100\]
\[
\begin{align*}
1 \\
2 \\
3 \\
4 \\
5 \\
6 \\
7
\end{align*}
\]
RISC-V Calling Conventions: Some useful facts to know

Arguments are passed in a0, a1, etc. with the first argument being in a0, the second in a1 and so on.

Return values are returned in a0 (and a1 if you have two return values).

The return address is stored in the ra register.

A and T registers can be used freely provided you don’t need the original value in them. If you do, must save them to the stack before making another procedure call.

Ra must be saved onto the stack if you are making nested procedure calls otherwise your original return address will be overwritten.

S registers and the SP must be returned in exactly the same state as they were when a function was called. This means all S registers must be saved onto the stack before modifying them. Prior to returning from a procedure, anything that was pushed onto the stack must be taken off the stack so that SP is restored to its original value.

To push N values onto the stack:
addi sp, sp, -X       // X = N*4
sw x1, 0(sp)         // value 1
sw x2, 4(sp)         // value 2
sw x3, 8(sp)         // …
sw x4, 0xC(sp)       // value N

To pop N values from the stack without restoring the values:
addi sp, sp, X

To pop N values from the stack and restore the values:
lw x1, 0(sp)
lw x2, 4(sp)
lw x3, 8(sp)
lw x4, 0xC(sp)
addi sp, sp, X

To make a procedure call you must save the return address in ra. This can be achieved using either:
jal ra, label

or

jal label    (this is a pseudoinstruction that gets translated into jal ra,label).

To return from a procedure call you must set the pc to the contents of ra. This can be achieved using either:
ret    or    jr ra   or    jalr ra
**Problem 2. RISC-V Calling Conventions and Stacks**

For each implementation of the function `funcA` defined below, identify and correct any errors that result in incorrect behavior of the function as well as any errors in the use of the RISC-V calling conventions. Assume that `funcA` is called from code that follows the RISC-V calling conventions. Also, assume that `funcA` is originally called with \( x > 0 \).

`funcA` is defined as:

```c
int funcA(int a, int b, int x) {
    if (x == 0) return b;
    else return x + funcA(a+b, b+2, x-1);
}
```

Implementation 1:

```
funcA:
  beqz a2, done
  addi sp, sp, -4
  sw ra, 0(sp)
  add a0, a0, a1
  addi a1, a1, 2
  addi a2, a2, -1
  jal ra, funcA
  add a0, a0, a2
  lw ra, 0(sp)
  addi sp, sp, 4
  ret

done:  mv a0, a1  // return b
  ret
```

Implementation 2:

```
funcA:
  beqz a2, done
  addi sp, sp, -8
  sw ra, 0(sp)
  sw a2, 4(sp)
  add a0, a0, a1
  addi a1, a1, 2
  addi a2, a2, -1
  jal ra, funcA
  add a0, a0, a2
  lw ra, 0(sp)
  lw a2, 4(sp)
  addi sp, sp, 8
  ret

done:  mv a0, a1  // return b
```
Implementation 3:

funcA:
  beqz a2, done
  addi t1, a1, 2
  addi t0, a0, a1  // t0 = a0 + a1
  addi s2, a2, -1
  mv t2, a2
  mv a0, t0
  mv a1, t1
  mv a2, s2
  jal ra, funcA
  add a0, a0, t2
  ret

done:
  mv a0, a1  // return b
  ret
  ret
Problem 3:
Stack Detective:

Assume that the code for the previous problem was written as shown below:

```c
int funcA(int a, int b, int x) {
    if (x == 0) return b;
    else return x + funcA(a+b, b+2, x-1)
}
```

```asm
.beqz a2, done
.addi sp, sp, -8
.sw ra, 0(sp)
.sw a2, 4(sp)
.add a0, a0, a1
.addi a1, a1, 2
.addi a2, a2, -1
.jal ra, funcA
.lw a2, 4(sp)
L1:    add a0, a0, a2
       lw ra, 0(sp)
       addi sp, sp, 8
       ret

done:   mv a0, a1       // return b
       ret
```

Assume execution was halted just as you are about to execute the instruction at label L1.

You are given the stack on the right at the time execution was halted.

A) What is the value of the empty cell in the stack trace above?

B)

<table>
<thead>
<tr>
<th>Original x</th>
<th>Original ra</th>
</tr>
</thead>
<tbody>
<tr>
<td>A = 0x1B0</td>
<td></td>
</tr>
<tr>
<td>A = 0x228</td>
<td></td>
</tr>
</tbody>
</table>
Caches:

Problem 4. Cache Hit Rates

The following questions ask you to evaluate alternative cache designs using a specific pattern of memory references. Each of the caches under consideration has a total capacity of 8 (4-byte) words, with one word stored in each cache line. The cache designs under consideration are:

DM: a direct-mapped cache.

S2: a 2-way set-associative cache with a least-recently-used (LRU) replacement policy.

The questions below present a repeating sequence of addresses (expressed as decimal numbers) for memory reads. After the access to the last address on the list, the sequence starts over with the first address on the list. Keep in mind that byte addressing is used; addresses of consecutive words in memory differ by 4. Each question asks for the hit ratio for each cache. Answer by considering the steady-state hit ratio, i.e., the percentage of memory references that hit in the cache after the sequence has been repeated many times. Your answer can be expressed as a percentage or a ratio.

<table>
<thead>
<tr>
<th>Repeating address sequence</th>
<th>Hit ratio for cache DM</th>
<th>Hit ratio for cache S2</th>
</tr>
</thead>
<tbody>
<tr>
<td>(A) 0, 8, 16, 24, 32</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(B) 0, 4, 8, 12, 16, 20, 24, 28, 32</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(C) 0, 4, 8, 12, 32, 36, 40, 44</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Problem 5. Memory Hierarchy and Cache Accesses

(A) (1 point)

Assuming the memory hierarchy configuration shown above which consists of a level 1 (L1) cache with access time of 4 ns and hit rate of 0.6, and a level 2 (L2) cache with access time of 8 ns and hit rate of 0.9, and a main memory whose access time is 100 ns, determine the AMAT (average memory access time) for this system.

AMAT (ns): ______________

(B) (1 point) True or False. A 4 line 2-way set associative cache always has a better hit rate than a direct mapped cache with 8 lines.

TRUE      FALSE

For parts C-G, use the 2-line 2-way set associative cache with a block size of 2, shown below.

(C) (1 point) Which address bits are used for the offset, cache line index, and tag comparison?

Address bits used as offset (including byte offset): A[ _____ : _____ ]

Address bits used as cache line index: A[ _____ : _____ ]

Address bits used for tag comparison: A[ _____ : _____ ]

(D) (1 point) Give the address corresponding to the data 0x23 (line 0, data word B1) or write CAN’T TELL if there’s insufficient information to determine its address.

CAN’T TELL of address of 0x23 data (0x): ______________

(E) (1 point) Would a store to address 0x71C hit or miss?

HIT      MISS
(F) (1 point) Suppose that the CPU executes a sw instruction writes the value 7 into location 0x71C. Fill in all changed entries for the cache in the unfilled cache below.

For parts G and H compute the line number and tag for each data access. Indicate if the access will result in a hit or a miss. Provide the value returned by the lw, or enter CAN’T TELL if you can’t tell from the information provided. (1 point each)
Problem 6: Hardware Implementation

We want to add the following instruction to the RISC-V ISA:

\texttt{lwr} r_d, r_s1, r_s2

The behavior of this new \texttt{lwr} instruction is that it adds the contents of source registers \( r_s1 \) and \( r_s2 \) and uses the result as the address for your load. In other words:

\[
\text{reg}[r_d] \leftarrow \text{mem}[\text{reg}[r_s1] + 4 \times \text{reg}[r_s2]] \\
\text{pc} \leftarrow \text{pc} + 4
\]

Suppose you were given a decoder that was fully functioning prior to the addition of this new instruction, and you were told that a new type \texttt{IType LWR} has been added for this instruction. Fill in the required changes to the Decode and Execute functions in order to support adding this instruction. We have included relevant portions of these functions below to simplify this task. Assume that the rest of the hardware implementation is the same as what you have seen in lecture and lab.

(A) (12 points) Fill in the required additions to the code below.

```verilog
typedef enum { OP, OPIMM, BRANCH, LUI, JAL, JALR, LOAD, STORE, GCD, AUIPC, LWR, Unsupported} IType deriving (Bits, Eq, FShow);

function DecodedInst decode(Bit#(32) inst);

    Bit#(7) opLWR = 7'b__________;

    DecodedInst dInst = ?;
    dInst.iType = Unsupported;
    case(opcode)
        ...
        opLoad: if (funct3 == fnLW) begin
            dInst = DecodedInst { rd: tagged Valid rd, rs1: r_s1, rs2: ?, imm: immI, brFunc: ?,
                                aluFunc: ?, iType: LOAD };
        end
        opLWR:

        endcase
    return dInst;
endfunction
```

Write your code below. Please add arrows to indicate where your code fits in the decode function.
function ExecInst execute(DecodedInst dInst, Word rVal1, Word rVal2, Word pc);
    let imm = dInst.imm;
    let brFunc = dInst.brFunc;
    let aluFunc = dInst.aluFunc;
    Word data = ?;
    Word nextPc = ?;
    Word addr = ?;
    case (dInst.iType)
        OP: begin data = alu(rVal1, rVal2, aluFunc); nextPc = pc+4; end
        OPIMM: begin data = alu(rVal1, imm, aluFunc); nextPc = pc+4; end
        BRANCH: begin nextPc=aluBr(rVal1,rVal2,brFunc)? pc+imm : pc+4; end
        LUI: begin data = imm; nextPc = pc+4; end
        JAL: begin data = pc+4; nextPc = pc+imm; end
        JALR: begin data = pc+4; nextPc = (rVal1+imm) & ~1; end
        LOAD: begin addr = rVal1+imm; nextPc = pc+4; end
        STORE: begin data = rVal2; addr = rVal1+imm; nextPc = pc+4; end
        AUIPC: begin data = pc+imm; nextPc = pc+4; end
        LWRI: begin data = pc+imm; nextPc = pc+4; end
    endcase
    ExecInst eInst = ?;
    eInst.iType = dInst.iType;
    eInst.rd = dInst.rd;
    eInst.data = data;
    eInst.addr = addr;
    eInst.nextPc = nextPc;
    return eInst;
endfunction

Write your code below. Please add arrows to indicate where your code fits in the `execute` function.

(B) (3 points) Could a store operation that calculated the address in the same manner as LWRI be added without major changes to the RISC-V ISA?

Yes                  No

(C) (3 points) If you answered Yes to B then provide a brief summary of the required changes. If you answered No to B then explain why not.

(D) Do we need to add an extra stage in our multi-cycle implementation to implement LWRI?